비교 연구에서 표본크기 결정

Sample size determination in comparative studies

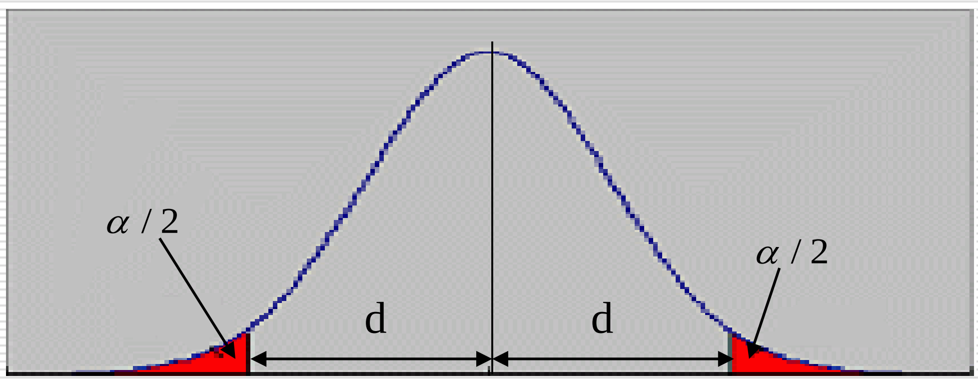
2018년 1학기

Contents

One sample case : introduction

1) Estimating the population proportion

cf. $E(p) = P,\quad Var(p) = P(1-P)/n$



P

$P - z_{1-\alpha/2}\sqrt{P(1-P)/n}$ $\qquad$ $P + z_{1-\alpha/2}\sqrt{P(1-P)/n}$

- The desired precision using absolute error $d = |\hat{P} - P|$

$$n = \frac{z_{1-\alpha/2}^{2} P(1-P)}{d^{2}} \quad \text{from} \quad d = |\hat{P} - P| = z_{1-\alpha/2}\sqrt{P(1-P)/n}$$

- The desired precision using relative error $\varepsilon = \dfrac{|\hat{\theta} - \theta|}{\theta}$

$$\varepsilon = \frac{|\hat{P} - P|}{P} = z_{1-\alpha/2}\frac{\sqrt{P(1-P)}}{\sqrt{n}\,P} = z_{1-\alpha/2}\frac{\sqrt{1-P}}{\sqrt{nP}}$$

$$n = \frac{z_{1-\alpha/2}^{2}(1-P)}{\varepsilon^{2} P}$$

- The desired precision using relative error (CV)

$$\varepsilon = \frac{|\hat{P} - P|}{P} = z_{1-\alpha/2}\frac{\sqrt{P(1-p)}\Big/\sqrt{n}}{P} = z_{1-\alpha/2}\frac{CV}{\sqrt{n}} = z_{1-\alpha/2} \times CV_{\bar{y}}$$

$$n = \frac{z^2_{1-\alpha/2} \times (CV)^2}{\varepsilon^2} = \frac{z^2_{1-\alpha/2} \times (1-P)}{\varepsilon^2 P}$$

- The desired precision using relative standard error

$$\varepsilon = z_{1-\alpha/2} \times CV_{\bar{y}} = z_{1-\alpha/2} \times CV\Big/\sqrt{n} \quad \text{where} \quad CV = \frac{S}{P} = \frac{\sqrt{P(1-P)}}{P} = \sqrt{(1-P)\Big/P}$$

$$n = \left(\frac{CV}{CV_{\bar{y}}}\right)^2$$

- Continuous study

n' and $CV'_{\bar{y}}$ are the sample size and CV of the past survey.

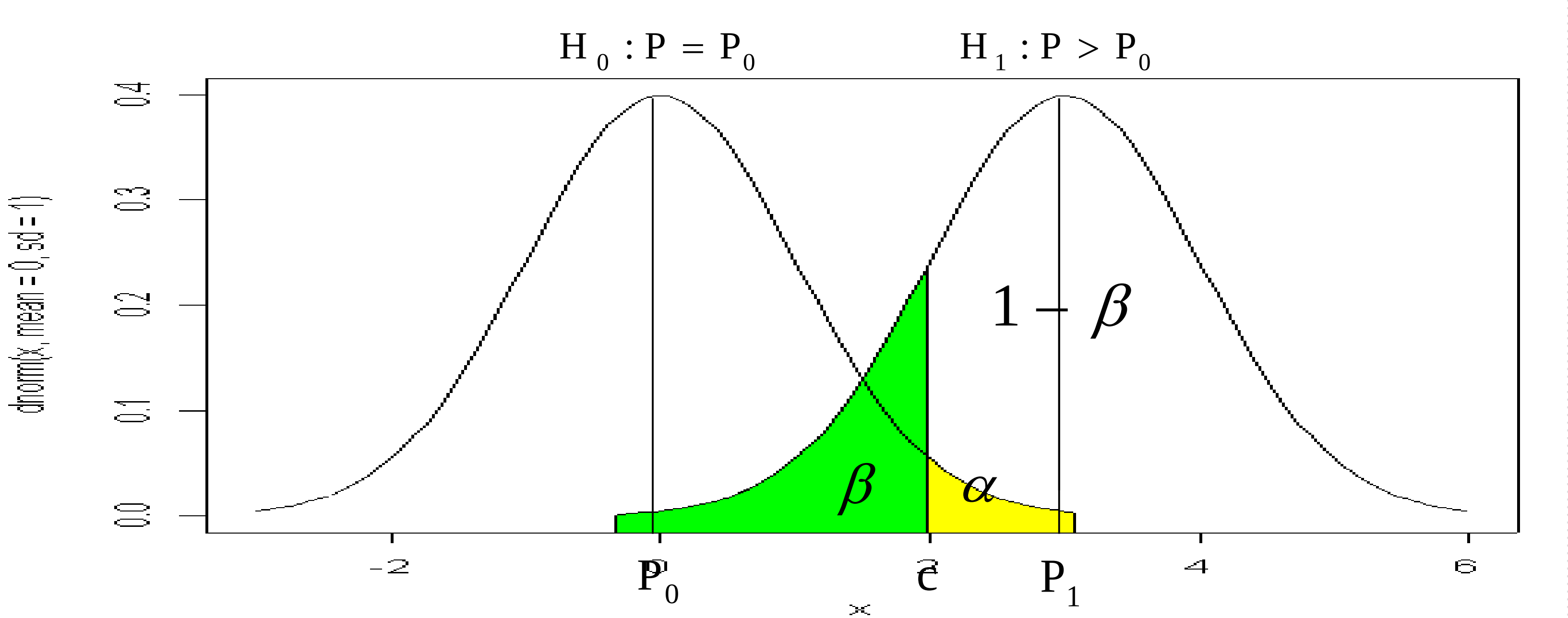
$$\text{n} = n' \times \left(\frac{CV'_{\bar{y}}}{CV_{\bar{y}}} \right)^2$$

- Without Replacement

$$n_0 = \left(\frac{z_{1-\alpha/2}^{\ 2} \times P(1-P)}{d^2} \right) \quad : \text{With Replacement}$$

$$\text{n} = \frac{n_0}{1 + n_0 / N}$$

2) Hypothesis testing for a single population proportion(1)



$$c = P_0 + z_{1-\alpha}\sqrt{P_0(1-P_0)/n} \quad \text{under} \quad H_0$$

$$c = P_1 - z_{1-\beta}\sqrt{P_1(1-P_1)/n} \quad \text{under} \quad H_1$$

$$c = P_0 + z_{1-\alpha}\sqrt{P_0(1-P_0)/n} = P_1 - z_{1-\beta}\sqrt{P_1(1-P_1)/n}$$

$$P_1 - P_0 = z_{1-\alpha}\sqrt{P_0(1-P_0)/n} + z_{1-\beta}\sqrt{P_1(1-P_1)/n}$$
$$= \left\{z_{1-\alpha}\sqrt{P_0(1-P_0)} + z_{1-\beta}\sqrt{P_1(1-P_1)}\right\} / \sqrt{n}$$

$$n = \frac{\left\{z_{1-\alpha}\sqrt{P_0(1-P_0)} + z_{1-\beta}\sqrt{P_1(1-P_1)}\right\}^2}{\left(P_1 - P_0\right)^2}$$

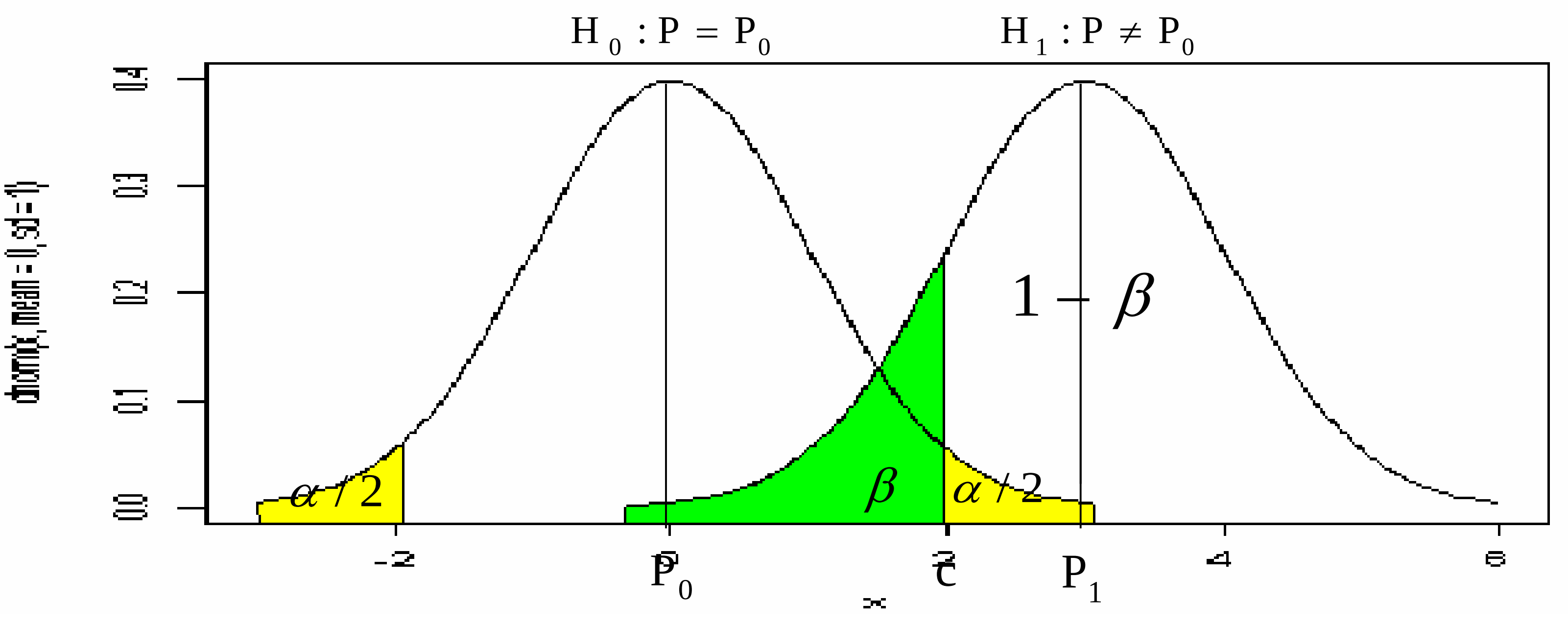
Example 1

- 출생 시 발병률이 1000명당 150명에게서 나타나는 질병이 있다고 가정하자. 어떤 의사는 임상수준이 향상되어 1000명당 100명으로 감소되는 경향을 보이고 있다고 주장하고 있다. 유의수준 0.05수준에서 귀무가설 $(H_0 : P = 0.15)$ 을 검정하려고 하는 데, 90%의 수준에서 감소된 발병률(검정결과)이 참이라고 확신하기를 원하고 있다. 이 때 필요한 표본크기는 얼마인가?

$$n = \frac{\left\{ z_{1-\alpha} \sqrt{P_0(1-P_0)} + z_{1-\beta} \sqrt{P_1(1-P_1)} \right\}^2}{(P_1 - P_0)^2}$$

$$= \frac{\left\{ 1.645 \sqrt{(0.15)(0.85)} + 1.282 \sqrt{(0.1)(0.9)} \right\}^2}{(0.1 - 0.15)^2} = 377.9$$

3) Hypothesis testing for a single population proportion(2)



$$c = P_0 + z_{1-\alpha/2}\sqrt{P_0(1-P_0)/n} \quad \text{under} \quad H_0$$

$$c = P_1 - z_{1-\beta}\sqrt{P_1(1-P_1)/n} \quad \text{under} \quad H_1$$

$$c = P_0 + z_{1-\alpha/2}\sqrt{P_0(1-P_0)/n} = P_1 - z_{1-\beta}\sqrt{P_1(1-P_1)/n}$$

$$\begin{aligned} P_1 - P_0 &= z_{1-\alpha/2}\sqrt{P_0(1-P_0)/n} + z_{1-\beta}\sqrt{P_1(1-P_1)/n} \\ &= \left\{z_{1-\alpha/2}\sqrt{P_0(1-P_0)} + z_{1-\beta}\sqrt{P_1(1-P_1)}\right\}/\sqrt{n} \end{aligned}$$

$$n = \frac{\left\{z_{1-\alpha/2}\sqrt{P_0(1-P_0)} + z_{1-\beta}\sqrt{P_1(1-P_1)}\right\}^2}{(P_1 - P_0)^2}$$

Example 2

- 어느 질병의 처치 성공률이 0.7이라고 알려져 있는 데, 동일한 성공률을 보장하는 새로운 처치법이 개발되었다. 담당전문의의 도움을 받을 수 없는 어느 병원에서 새로운 처치법을 환자에게 사용하려고 한다. 유의수준 0.05 수준에서 두 가설 $(H_0 : P = 0.70)$과 $(H_0 : P \neq 0.70)$ 에 대해 검정하려고 하는 데, 90%의 수준에서 기존의 처치와 10%p이상 차이가 나지 않기를 원한다. 이 조사를 수행하는 데 필요한 표본크기는 얼마인가?

$$\text{since} \quad |P_0 - P_1| \leq 0.1, \quad P_1 = 0.8 \quad \text{or} \quad 0.6$$

$$n_2 = \frac{\left\{ z_{1-\alpha/2}\sqrt{P_0(1-P_0)} + z_{1-\beta}\sqrt{P_1(1-P_1)} \right\}^2}{\left(P_1 - P_0\right)^2}$$

$$= \frac{\left\{ 1.96\sqrt{(0.7)(0.3)} + 1.282\sqrt{(0.6)(0.4)} \right\}^2}{(0.7-0.8)^2} = 232.94$$

$$n_1 = \frac{\left\{ z_{1-\alpha/2}\sqrt{P_0(1-P_0)} + z_{1-\beta}\sqrt{P_1(1-P_1)} \right\}^2}{\left(P_1 - P_0\right)^2}$$

$$= \frac{\left\{ 1.96\sqrt{(0.7)(0.3)} + 1.282\sqrt{(0.8)(0.2)} \right\}^2}{(0.7-0.8)^2} = 199.09$$

$$\therefore \quad n = \max(n_1, n_2) = 233$$

Two samples case

1) Estimating the difference between two proportions (or the risk difference)

$$\text{cf.}\quad E(p_1 - p_2) = P_1 - P_2,$$

$$\mathrm{Var}(p_1 - p_2) = P_1(1 - P_1)/n_1 + P_2(1 - P_2)/n_2$$

$$\text{if}\quad n_1 = n_2 = n,\ \text{then}\quad \mathrm{Var}(p_1 - p_2) = \left(1/n\right)\{P_1(1 - P_1) + P_2(1 - P_2)\}$$

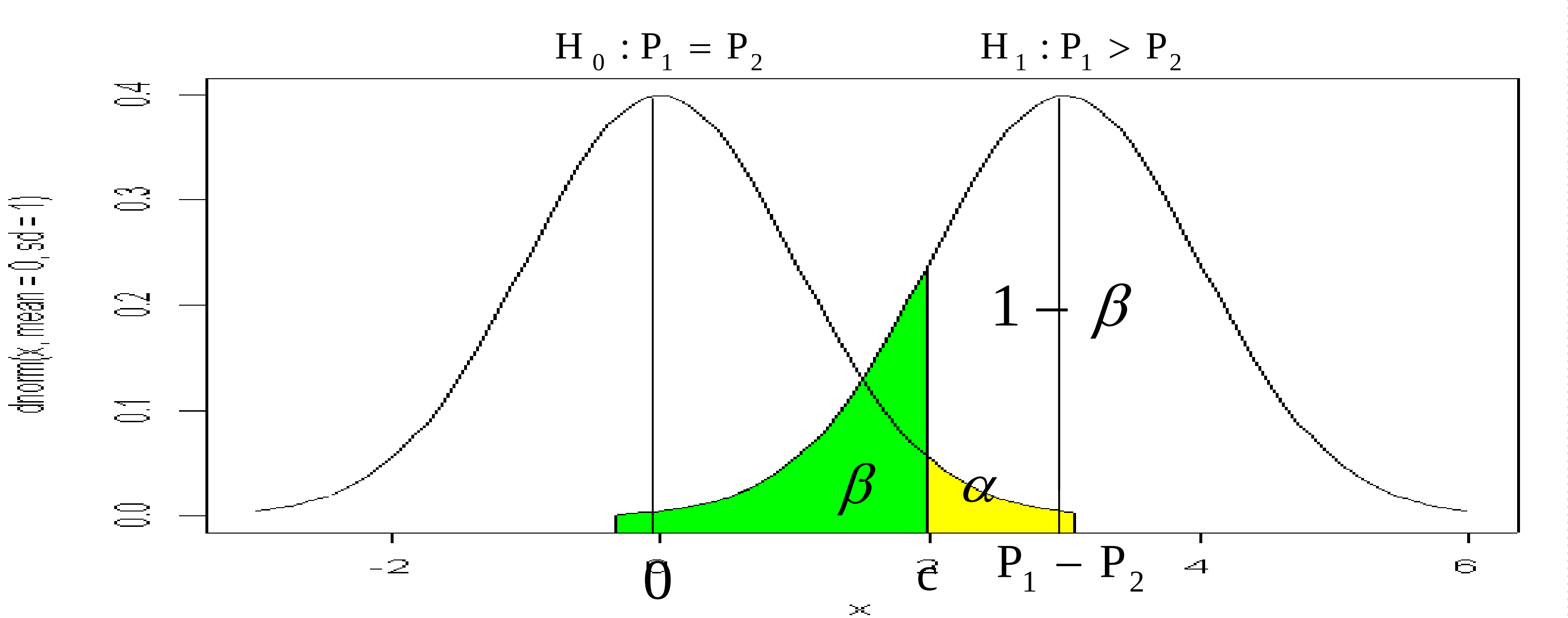
- The desired precision using absolute error and sample size

$$d = z_{1-\alpha/2}\sqrt{(1/n)\ \{P_1(1 - P_1) + P_2(1 - P_2)\}}$$

$$n = \frac{z_{1-\alpha/2}^{2}\{P_1(1 - P_1) + P_2(1 - P_2)\}}{d^2}$$

- if $n_2 = k\, n_1$, then $n = \dfrac{z_{1-\alpha/2}^{2}\{k \cdot P_1(1 - P_1) + P_2(1 - P_2)\}}{k\, d^2}$

2) Hypothesis testing for two population proportions(1)



$$\mathrm{Var}(P_1 - P_2) = P_1(1-P_1)/n_1 + P_2(1-P_2)/n_2 \quad \text{under} \quad H_0$$

$$= P(1-P)(1/n_1 + 1/n_2) \quad \text{if} \quad P_1 = P_2 = P$$

$$= 2\left\{P(1-P)/n\right\} \quad \text{if} \quad n_1 = n_2 = n$$

cf. $\hat{P} = \bar{p} = \left(p_1 + p_2\right)/2$

● Sample size

$$c = 0 + z_{1-\alpha}\sqrt{2\bar{P}(1-\bar{P})/n} \quad \text{under } H_0, \quad \text{where } \bar{P} = (P_1 + P_2)/2$$

$$c = (P_1 - P_2) - z_{1-\beta}\sqrt{(1/n)\left\{P_1(1-P_1) + P_2(1-P_2)\right\}} \quad \text{under } H_1,$$

assuming $n_1 = n_2 = n$

$$n = \frac{\left\{z_{1-\alpha}\sqrt{2\bar{P}(1-\bar{P})} + z_{1-\beta}\sqrt{P_1(1-P_1) + P_2(1-P_2)}\right\}^2}{(P_1 - P_2)^2}$$

Example 3

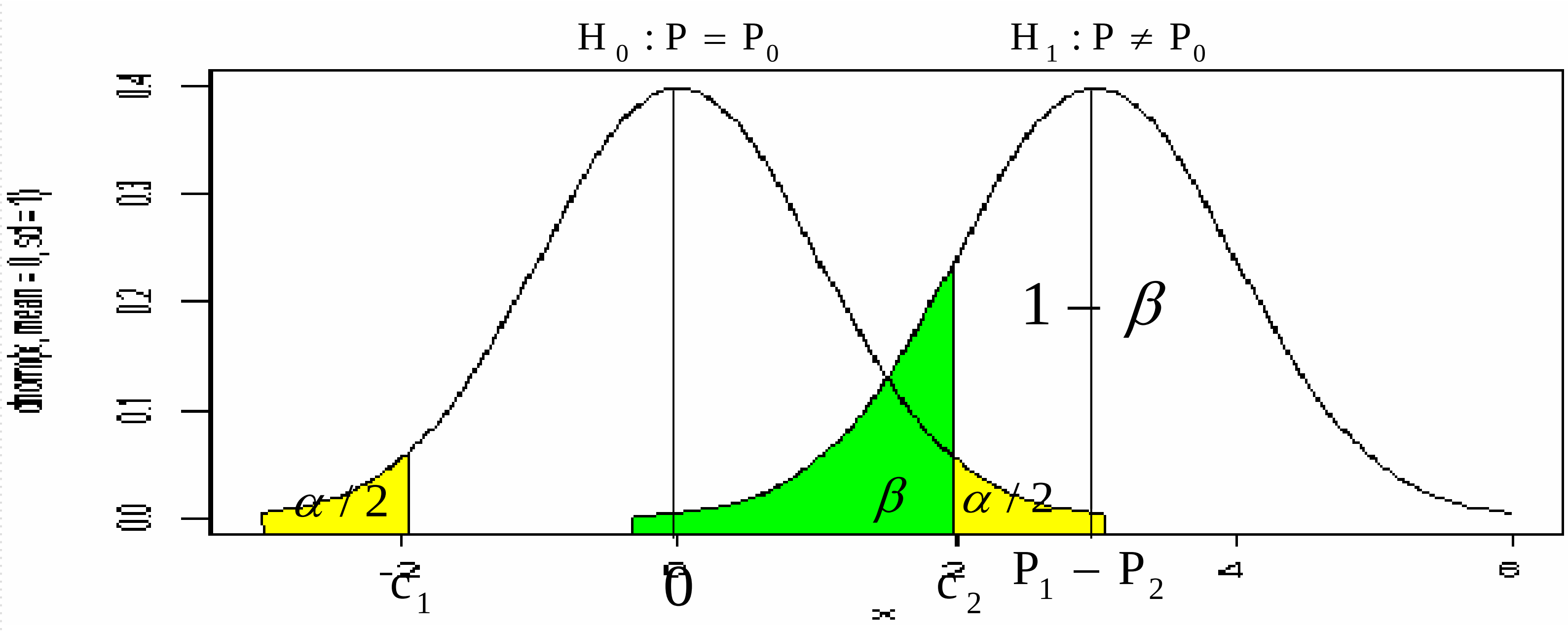
- 충치발병률이 0.8인 A지역과 0.6인 B지역이 있다고 가정하자. 두 지역의 발병률차이가 10%인지 유의수준 10%에서 검정하려고 하는 데, 80%의 수준에서 이 차이가 실제 차이(검정결과)로 확신하고자 한다. 이 조사를 위한 필요한 표본크기는 각각 얼마로 결정하면 되는가?

$$\overline{P} = (P_1 + P_2)/2 = 0.7$$

$$n = \frac{\left\{ z_{1-\alpha}\sqrt{2\,\overline{P}(1-\overline{P})} + z_{1-\beta}\sqrt{P_1(1-P_1) + P_2(1-P_2)} \right\}^2}{(P_1 - P_2)^2}$$

$$= \frac{\left\{ 1.282\sqrt{2(0.7)(0.3)} + 0.842\sqrt{(0.8)(0.2) + (0.6)(0.4)} \right\}^2}{(0.8 - 0.6)^2} = 46.47$$

3) Hypothesis testing for two population proportions(2)



$$c_2 = 0 + z_{1-\alpha/2}\sqrt{2\bar{P}(1-\bar{P})/n} \quad \text{under} \quad H_0, \quad \text{where} \quad \bar{P} = (P_1 + P_2)/2$$

$$c_2 = (P_1 - P_2) - z_{1-\beta}\sqrt{(1/n)\ \{P_1(1-P_1) + P_2(1-P_2)\}} \quad \text{under} \quad H_1,\ n_1 = n_2 = n$$

- Sample size

$$n = \frac{\left\{ z_{1-\alpha/2} \sqrt{2\,\overline{P}(1-\overline{P})} + z_{1-\beta} \sqrt{P_1(1-P_1) + P_2(1-P_2)} \right\}^2}{\left(P_1 - P_2\right)^2}$$

Example 4

- 신경성 전염질환을 연구하는 역학연구자는 질환을 앓고 있는 그룹(실험군)과 무관한 그룹(대조군)을 비교하고자 50명을 각각 랜덤하게 표집하여 본 결과, 특정 화학약품을 취급하는 공장에 종사하는 비율이 실험군에서는 0.6, 대조군에서는 0.5인 것으로 나타났다. 예비조사에서 얻은 결과는 이 지역에 거주하는 인구 중 화학약품을 취급하는 공장에 근무하는 비율과 같다고 가정할 때, 이 차이가 유의한 지 유의수준 5%에서 검정하려고 하는데, 90%의 수준에서 두 그룹간 차이(검정결과)를 실제 차이로 인정하고자 한다. 이 연구를 위해 추가하여 표집해야 하는 표본크기는 얼마인가?

$$P_1 = 0.6,\quad P_2 = 0.5,\quad \alpha = .05,\quad \beta = .1$$

$$n = \frac{\left\{ z_{1-\alpha/2}\sqrt{2\,\bar{P}(1-\bar{P})} + z_{1-\beta}\sqrt{P_1(1-P_1) + P_2(1-P_2)} \right\}^2}{\left(P_1 - P_2\right)^2}$$

$$= \frac{\left\{ 1.96\sqrt{2(0.55)(0.45)} + 1.282\sqrt{(0.6)(0.4)+(0.5)(0.5)} \right\}^2}{(0.6-0.5)^2} = 518.19$$

[Review]
Epidemiologic Study Designs

1. Some basic concepts

- The cohort or follow-up study
  - 대상(개인)을 특정한 특성의 존재 여부에 따른 그룹화하여 관심을 두고 있는 발생결과와의 관련성 여부를 규명하는 역학 조사연구방법의 분야(전향적 연구 방법)
    - 특정 특성 : 흡연여부 등, the exposure variable
    - 발생 결과 : 질병의 존재 여부 등

- A case-control design/study
  - 발생 결과를 토대로 사례군(a case group)과 대조군(a control group)으로 분류하여 각 집단에 대해 과거 특정한 위험인자(a risk factor, 특성)의 존재 여부를 조사하는 역학 조사연구방법(후향적 연구 방법)

A prevalence study(유병율 연구)

- 모집단으로부터 랜덤 표집된 표본에서 특정 시점에서 측정된 특정 특성이나 위험인자의 존재나 발병에 대한 비율을 연구하는 방법
- (cf.) Incidence(발생율, 발병율) : 특정 기간 동안 모집단에서 새로이 발현된 혹은 나타난 병의 비율을 의미
- 일반적으로 유병율이 발병율보다 같거나 큰 값을 제공

2. The relative risk(RR) and odds ratio(OR)

- 역학 연구 분야의 코호트 연구나 사례-대조 연구에서의 중요한 모수

		Exposure		
		Present E	Absent $\bar{E}$	Total
Disease	Present D	a	b	n_1
	Absent $\bar{D}$	c	d	n_2
	Total	m_1	m_2	n

$$P_1 = \Pr(D \mid E),\ \ 1 - P_1 = \Pr(\bar{D} \mid E),\ \ P_2 = \Pr(D \mid \bar{E}),\ 1 - P_2 = \Pr(\bar{D} \mid \bar{E})$$

$$P_1' = \Pr(E \mid D),\ \ 1 - P_1' = \Pr(\bar{E} \mid D),\ \ P_2' = \Pr(E \mid \bar{D}),\ 1 - P_2' = \Pr(\bar{E} \mid \bar{D})$$

Relative Risk (RR, 상대 위험율)

- 표현식과 추정량

$$RR = \frac{\text{실험군에서의 \quad 위험률}}{\text{대조군에서의 \quad 위험률}} = \frac{\Pr(D \mid E)}{\Pr(D \mid \bar{E})} = \frac{P_1}{P_2}$$

$$R\hat{R} = \frac{a / m_1}{b / m_2}$$

- 코호트 연구나 사례군과 대조군이 미리 정해진 후 그 결과를 관찰하는 임상시험 연구에서 계산 가능
- 사례-대조 연구에서는 계산 불가능

□ Odds Ratio (OR, 오즈비)

- 표현식 : 코호트 연구

$$OR = \frac{\text{실험군에서의 오즈비}}{\text{대조군에서의 오즈비}} = \frac{O_1}{O_2} = \frac{\Pr(D \mid E)/\Pr(\overline{D} \mid E)}{\Pr(D \mid \overline{E})/\Pr(\overline{D} \mid \overline{E})} = \frac{P_1/(1-P_1)}{P_2/(1-P_2)}$$

- 표현식 : 사례-대조연구

$$OR = \frac{\text{실험군에서의 오즈비}}{\text{대조군에서의 오즈비}} = \frac{O_1}{O_2} = \frac{\Pr(E \mid D)/\Pr(\overline{E} \mid D)}{\Pr(E \mid \overline{D})/\Pr(\overline{E} \mid \overline{D})} = \frac{P_1'/(1-P_1')}{P_2'/(1-P_2')}$$

- 추정량 : the cross product ratio

$$O\hat{R} = \frac{\left(a/m_1\right)\Big/\left(c/m_1\right)}{\left(b/m_2\right)\Big/\left(d/m_2\right)} \; or \; \frac{\left(a/n_1\right)\Big/\left(b/n_1\right)}{\left(c/n_2\right)\Big/\left(d/n_2\right)} = \frac{ad}{bc}$$

OR 의 표본분포

- OR의 특성
  - OR의 범위 $0 \le OR \le \infty$
  - $OR = 1$ 의 의미 : 위험율이 노출그룹과 비노출그룹, 혹은 사례군과 대조군의 오즈비가 동일함을 의미
  - $\hat{OR}$ 의 표본 분포 특성
    - 비정규분포, 꼬리가 큰 값으로 긴 강한 양의 왜도를 갖는 분포
    - 자연로그변환 $\ln(\hat{OR})$ 을 통한 분포 특성 개선 : 보다 근사정규분포
    - 신뢰구간(표준정규분포 이용)은 비대칭
- $\ln(\hat{OR})$ 의 분산과 추정량

$$Var[\ln(\hat{OR})] \approx \left(\frac{1}{n_1}\right)\frac{1}{P_1'(1-P_1')} + \left(\frac{1}{n_2}\right)\frac{1}{P_2'(1-P_2')}$$

$$\hat{V}ar[\ln(\hat{OR})] = \frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d}$$

- $H_0 : OR = 1 \;\; vs \;\; H_1 : OR \ne 1$ 에 대한 검정 : 2X2 카이제곱검정

RR 의 표본분포

- $R\hat{R}$ 의 표본분포 특성
  - 매우 큰 표본에서는 RR은 근사정규분포
  - 역학연구의 표본크기 수준에서는 정규분포 형태가 아니며, 양의 왜도를 나타내는 분포
    - 자연로그 변환 $\ln(R\hat{R})$ 을 통해 보다 대칭적인 분포 형태를 취하게 되며, 정규분포로 근사
- 추정량 : 코호트 연구에서 계산 가능하므로 $R\hat{R} = p_1/p_2$
- $\ln(R\hat{R})$ 의 분산과 추정량

$$Var[\ln(R\hat{R})] = Var[\ln(p_1) - \ln(p_2)] = Var[\ln(p_1)] + Var[\ln(p_2)]$$

$$\hat{V}ar[\ln(R\hat{R})] = \left(\frac{1}{m_1}\right)\frac{1-p_1}{p_1} + \left(\frac{1}{m_2}\right)\frac{1-p_2}{p_2} \quad \text{since} \quad Var[\ln(p)] = \left(\frac{1}{n}\right)\frac{1-p}{p}$$

- 신뢰구간은 OR과 동일한 특성을 보임

3. Screening test for disease prevalence

- 초음파나 X-ray, PAP test등의 간이검사 결과를 통해 발병 진단 시 오류 발생
- 진단법의 평가 결과 (H_0 : 환자)

		결과/환자의 실제 상황		
		환자	정상	Total
간이검사 진단 결과	발병 가능/예측 (양성, positive)	a True positives	b False positives	n_1
	정상 예측 (음성, negative)	c False negatives	d True negatives	n_2
	Total	m_1	m_2	n

□ 진단법의 평가 척도 : 민감도와 특이도

1) **민감도(sensitivity)** : 환자를 양성으로 판정하는 비율

$$\text{true positive rate} = a/m_1$$

2) False negative rate $\text{false negative rate} = c/m_1$

➢ 검정의 1종의 오류와 동일한 의미

3) False positive rate $\text{false positive rate} = b/m_2$

➢ 검정의 2종의 오류와 동일한 의미

4) **특이도(specificity)** : 정상을 음성으로 판정하는 비율

➢ 간이진단법의 진단 능력을 보여주는 척도 = 검정력

$$\text{true negative rate} = d/m_2$$

□ 진단법의 평가 : 예측

- 현실에서는 실제 환자 수와 정상인 수를 모르므로 민감도와 특이도를 정확하게 계산할 수 없는 상황

- 진단 결과에 대한 예측 : 병의 유병률에 의존

1) 양성예측(the predictive value of a positive) a/n_1

2) 음성예측 (the predictive value of a positive) d/n_2

- 예측도(the predictive value of the screening test)
  : Bayes 정리 이용

1) 양성예측도

$$\Pr(\text{환자} \mid \text{양성}) = \frac{\Pr(\text{환자})\Pr(\text{양성} \mid \text{환자})}{\Pr(\text{환자})\Pr(\text{양성} \mid \text{환자}) + \Pr(\text{정상})\Pr(\text{양성} \mid \text{정상})} = \frac{\text{유병률} \times \text{민감도}}{\text{유병률} \times \text{민감도} + (1-\text{유병률})(1-\text{특이도})}$$

2) 음성예측도

$$\Pr(\text{정상} \mid \text{음성}) = \frac{\Pr(\text{정상})\Pr(\text{음성} \mid \text{정상})}{\Pr(\text{정상})\Pr(\text{음성} \mid \text{정상}) + \Pr(\text{환자})\Pr(\text{음성} \mid \text{환자})} = \frac{(1-\text{유병률}) \times \text{특이도}}{(1-\text{유병률}) \times \text{특이도} + \text{유병률} \times (1-\text{민감도})}$$

Sample size for case-control studies

1) Estimating the odds ratio with stated precision ε

- $\ln(\hat{OR})$ 의 표본분포 특성
  - ✓ OR의 표본분포 :
    - 우측으로 꼬리(양의 왜도)가 긴 비대칭 분포
    - 상당히 큰 표본 규모일 때 정규분포와 근사
  - ✓ 표본크기에 영향을 받는 OR의 표본분포보다 정규분포와 더 근사
  - ✓ 오즈비 추론에 널리 이용

- $\ln(\hat{OR})$ 의 특성 : 사례-대조 연구의 경우
  - ✓ 현재 질병 존재 여부에 따른 특성의 발현(the exposure) 확률로 정의
  - ✓ $\ln(\hat{OR})$ 의 분산 : $n_1 = n_2 = n$

$$\mathrm{Var}[\ln(\hat{OR})] \approx \left(\frac{1}{n_1}\right)\frac{1}{P_1'(1-P_1')} + \left(\frac{1}{n_2}\right)\frac{1}{P_2'(1-P_2')} = \left(\frac{1}{n}\right)\left(\frac{1}{P_1'} + \frac{1}{1-P_1'}\right) + \left(\frac{1}{n}\right)\left(\frac{1}{P_2'} + \frac{1}{1-P_2'}\right)$$

$$\hat{V}ar[\ln(\hat{OR})] = \frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d}$$

[Review]

$H_0 : OR = 1$ vs $H_1 : OR \neq 1$ 에 대한 검정 $\chi^2(1)$

□ 표본크기 결정을 위한 가정

- $OR > 1$ 의 가정 : 특성의 존재(exposed)와 비존재(unexposed)를 교체하여 $OR > 1$ 이 되도록 조정
- $w = OR - O\hat{R}_L$ 를 기준으로 표본크기 결정 : $O\hat{R}_U - OR$ 를 기준으로 결정하면 지나치게 큰 표본크기로 계산됨

  (여기서 $O\hat{R}_L$ 과 $O\hat{R}_U$ 은 신뢰구간의 하한과 상한을 의미)

  ✓ ε 과 w 의 관계 $w = \varepsilon \cdot OR$, where $\varepsilon = |O\hat{R} - OR| / OR$

□ 표본크기 결정

① w 와 신뢰구간과의 관계

$$w = \varepsilon \cdot OR = e^{\ln(OR)} - e^{\ln(OR) - z \cdot SE[\ln(OR)]} = OR - OR_L$$

$$= OR - OR \cdot e^{-z \cdot SE[\ln(OR)]}$$

② ε 의 표현 $\varepsilon = 1 - e^{-z \cdot SE[\ln(OR)]}$

③ 자연로그변환 $1 - \varepsilon = e^{-z \cdot SE[\ln(OR)]} \Rightarrow \ln(1-\varepsilon) = -z_{1-\alpha/2} \cdot SE[\ln(OR)]$

$$= -z_{1-\alpha/2} \cdot \sqrt{\left(1/n\right)\left(1/[P_1'(1-P_1')] + 1/[P_2'(1-P_2')]\right)}$$

$$\therefore \quad n = \frac{z_{1-\alpha/2}^2 \cdot \left(1/[P_1'(1-P_1')] + 1/[P_2'(1-P_2')]\right)}{[\ln(1-\varepsilon)]^2}$$

Example 5

- 표본크기 결정을 위한 모수 P_1', P_2', OR 의 관계(사례군 추정)
  - 두 모수(대조군의 발현율, OR)에 대한 정보가 존재한다고 가정

$$P_1' = \frac{OR \cdot P_2'}{OR \cdot P_2' + (1 - P_2')}$$

- 대조군의 특성 발현율(the exposure rate)이 0.3이며 오즈비가 2 정도로 예상되는 모집단에서 95%신뢰수준에서 모집단 오즈비의 25%이내 결과를 얻기 위한 사례-대조연구에 필요한 표본크기는 각각 얼마인가?

$$P_1' = \frac{OR \cdot P_2'}{OR \cdot P_2' + (1 - P_2')} = \frac{2 \cdot 0.3}{2 \cdot 0.3 + 0(1 - 0.3)} = 0.46$$

$$\therefore n = \frac{z_{1-\alpha/2}^2 \cdot \left(1 \Big/ [P_1'(1 - P_1')] + 1 \Big/ [P_2'(1 - P_2')] \right)}{[\ln(1 - \varepsilon)]^2} = \frac{1.96^2 \left[1 \Big/ (0.46 \times 0.54) + 1 \Big/ (0.3 \times 0.7) \right]}{[\ln(1 - 0.25)]^2} = 407.91$$

2) Sample size for hypothesis testing of the odds ratio

오즈비의 가설검정에 대한 관점

- 사례-대조연구에서 오즈비 검정에 대한 문제는 두 그룹의 특성 발현율(the proportion exposure)이 같은 지를 검정하는 것과 동일 $H_0 : OR = 1$

표본크기 결정

1) 사례군의 특성 발현율 추정

$$P_1' = \frac{OR \cdot P_2'}{OR \cdot P_2' + (1 - P_2')}$$

2) 가설 설정

$$H_0 : P_1' = P_2' \quad \text{vs} \quad H_1 : P_1' \neq P_2'$$

$$\text{cf.} \quad H_0 : OR = 1 \quad \text{vs} \quad H_1 : OR \neq 1$$

3) 표본크기 결정 : 두 모비율 차이의 과정과 동일

① 대조군의 특성 발현율에 대한 모집단 정보가 확실한 경우

$$n = \frac{\left\{ z_{1-\alpha/2}\sqrt{2\,P_2'(1-P_2')} + z_{1-\beta}\sqrt{P_1'(1-P_1') + P_2'(1-P_2')} \right\}^2}{\left(P_1' - P_2'\right)^2}$$

② 두 그룹의 특성 발현율에 대한 정보가 없거나 불확신 하다면,

$$n = \frac{\left\{ z_{1-\alpha/2}\sqrt{2\,\overline{P}'(1-\overline{P}')} + z_{1-\beta}\sqrt{P_1'(1-P_1') + P_2'(1-P_2')} \right\}^2}{\left(P_1' - P_2'\right)^2}$$

$$\text{where} \quad \overline{P}' = \left(P_1' + P_2'\right)/\,2$$

Example 6

- 유아의 BCG 백신 효능에 대한 연구를 위해 유아기에 결핵 접종을 한 그룹과 하지 않은 그룹을 비교하고자 한다. 대조군에서의 BCG 접종율이 약 30%라고 할 때, 유의수준 5%에서 두 그룹의 오즈비 차이가 존재하는 지를 검정하며, 검정 결과를 80% 수준에서 확신(두 그룹의 접종율로 인지)하고자 한다. 오즈비가 2정도이면 두 그룹간 의미있는 차이로 고려한다고 할 때, 이 사례-대조 연구를 위해 각 그룹마다 어느 정도의 표본이 필요한가?

$$n = \frac{\left\{ z_{1-\alpha/2}\sqrt{2P_2'(1-P_2')} + z_{1-\beta}\sqrt{P_1'(1-P_1') + P_2'(1-P_2')} \right\}^2}{\left(P_1' - P_2'\right)^2} \quad \text{where} \quad P_1' = \frac{2 \cdot 0.3}{(2 \cdot 0.3 + 0.7)} = 0.4615$$

$$= \frac{\left(1.96\sqrt{2 \times 0.3 \times 0.7} + 0.842 \times \sqrt{0.4615 \times 0.5385 + 0.3 \times 0.7}\right)^2}{(0.4615 - 0.3)^2} = 129.79$$

$$n = \frac{\left\{ z_{1-\alpha/2}\sqrt{2\bar{P}'(1-\bar{P}')} + z_{1-\beta}\sqrt{P_1'(1-P_1') + P_2'(1-P_2')} \right\}^2}{\left(P_1' - P_2'\right)^2} \quad \text{where} \quad \bar{P}' = \frac{(P_1' + P_2')}{2} = \frac{(0.3 + 0.46)}{2} = 0.38$$

$$= \frac{\left(1.96\sqrt{2 \times 0.38 \times 0.62} + 0.842 \times \sqrt{0.4615 \times 0.5385 + 0.3 \times 0.7}\right)^2}{(0.4615 - 0.3)^2} = 140.69$$

Sample size determination for Cohort studies

- 모집단 상대위험율(RR)의 상대오차 ε 이내 추정 결과를 얻기 위한 표본크기 결정
  - 오즈비(OR)와 동일 $w = \varepsilon \cdot RR$, where $\varepsilon = |\hat{RR} - RR| / RR$

$$w = \varepsilon \cdot RR = e^{\ln(RR)} - e^{\ln(RR) - z \cdot SE[\ln(RR)]} = RR - RR_L$$

$$= RR - RR \cdot e^{-z \cdot SE[\ln(RR)]}$$

$$\Rightarrow \varepsilon = 1 - e^{-z \cdot SE[\ln(RR)]} \Rightarrow 1 - \varepsilon = -e^{-z \cdot SE[\ln(RR)]}$$

$$\Rightarrow \ln(1-\varepsilon) = -z_{1-\alpha/2} \cdot SE[\ln(RR)]$$

$$= -z_{1-\alpha/2} \cdot \sqrt{\left(\frac{1}{m}\right)\left(\frac{(1-P_1)}{P_1} + \frac{(1-P_2)}{P_2}\right)}$$

$$\therefore m = \frac{z_{1-\alpha/2}^2 \cdot \left(\frac{(1-P_1)}{P_1} + \frac{(1-P_2)}{P_2}\right)}{[\ln(1-\varepsilon)]^2}$$

Example 7

- 코호트 연구의 비발현그룹(the unexposure group)의 20%가 환자라고 하자. 발현 그룹과 비발현그룹의 상대위험이 약 1.75배 정도인 모집단에서 모집단 상대위험율의 10%범위 내에서 상대위험율(RR)을 95% 신뢰수준으로 추정하려고 하는 코호트 연구에서 각 그룹마다 필요한 표본크기의 규모는 어느 정도인가?

$$P_2 = 0.2 \quad \text{and} \quad P_1 = (RR)P_2 = 0.35$$

$$\therefore m = \frac{z_{1-\alpha/2}^2 \cdot \left(\frac{(1-P_1)}{P_1} + \frac{(1-P_2)}{P_2} \right)}{\left[\ln(1-\varepsilon)\right]^2} = \frac{1.96^2 \left(\frac{0.65}{0.35} + \frac{0.8}{0.2} \right)}{\left(\ln(1-0.1)\right)^2} = 2026.95$$

1) Sample size for Hypothesis testing of the population relative risk

- 기본적인 과정은 오즈비(OR)와 동일
  - 검정은 두 모비율 차이 검정과 내용면에서 동일
- 기본 가설 : 두 그룹의 상대 위험(질병 유병)은 동일하다 $H_0 : \mathrm{RR} = 1$ or $H_0 : \mathrm{P}_1 = \mathrm{P}_2$
  - 대립가설 : $H_1 : \mathrm{RR} > 1$ $(\text{or } H_1 : \mathrm{RR} < 1)$ or $H_1 : \mathrm{RR} \neq 1$

    $H_1 : \mathrm{P}_1 > \mathrm{P}_2$ $(\text{or } H_1 : \mathrm{P}_1 < \mathrm{P}_2)$ or $H_1 : \mathrm{P}_1 \neq \mathrm{P}_2$
- **표본크기 결정 : 양측검정의 경우**

$$n = \frac{\left\{ z_{1-\alpha/2}\sqrt{2\,\mathrm{P}(1-\mathrm{P})} + z_{1-\beta}\sqrt{\mathrm{P}_1(1-\mathrm{P}_1) + \mathrm{P}_2(1-\mathrm{P}_2)} \right\}^2}{(\mathrm{P}_1 - \mathrm{P}_2)^2}$$

where $\mathrm{P} = (\mathrm{P}_1 + \mathrm{P}_2)/2 = \mathrm{P}_2(RR+1)/2$, $\mathrm{P}_1 = RR \cdot \mathrm{P}_2$, $0 < RR < 1/\mathrm{P}_2$

Example 8

- 암을 치료하는 두 가지 치료법이 있는 데 환자는 처치 A와 B의 치료를 랜덤하게 선택하여 5년 동안 치료 받는다고 가정하자. 비특성 그룹(특성 비발현 그룹)의 유병율이 $P_2 = 0.35$ 이고, 상대위험이 $RR = 0.5$ 인 모집단에서 유의수준 5%수준에서 두 그룹의 상대위험이 동일한지 여부 $H_0 : RR = 1$ vs $H_0 : RR \neq 1$ 에 대한 검정을 실시하고자 하며, 이 때 검정결과를 90% 신뢰수준에서 확신하려고 한다. 이 연구를 위해 필요한 표본크기는 그룹별로 각각 어느 정도인가?

$$P_1 = RR \cdot P_2 = 0.5 \times 0.35 = 0.175 \quad , \quad P = P_2(RR + 1)/2 = 0.2625$$

$$n = \frac{\left\{ z_{1-\alpha/2}\sqrt{2\,P(1-P)} + z_{1-\beta}\sqrt{P_1(1-P_1) + P_2(1-P_2)} \right\}^2}{\left(P_1 - P_2\right)^2}$$

$$= \frac{\left\{ 1.96\sqrt{2 \cdot (0.2625)(0.7375)} + 1.282\sqrt{0.175 \cdot 0.825 + 0.35 \cdot 0.65} \right\}^2}{\left(0.175 - 0.35\right)^2} = 130.79$$

The incidence rate

- 많은 역학연구에서 특정시점에 위험에 노출된 전체 대상보다는 관심의 질병(혹은 특성)이 관찰된 대상인 새로이 나타난 질병 발생자 수를 측정하고자 한다.
  - 질병 발생률(the incidence rate)
    - 대상이 노출된 관측 기간 중 발생된 환자 수의 비로 측정
    - 질병률(the force of morbidity), 위험률(the hazard)

- **질병 발생률(발병률 or 위험함수)의 측정**
  - 발병률은 포아송분포 모수 λ 형태를 따름
  - 예 : 5년 동안 5명의 환자를 추적하여 보니 2년 이후에 1명, 3년 이후에 1명이 발병하였다. 연간 (평균)발병률은?

$$\hat{\lambda} = \text{발생환자수} \Big/ \text{개인별 관측기간동안 노출된 대상합} = 2 \Big/ (5+5+5+2+3) = 0.1$$

1) Sample size determination for the incidence rate studies(1)

- 위험(혹은 발병률)의 추정량 $\hat{\lambda} = d/F$ d = 관심사건의 수 , F = 총 추적 기간
  - 발병밀도(the incidence density)는 지수 생존함수 조건하의 위험함수 추정량과 동일
  - 표본크기 결정에 이용
  - 연구기간 동안 위험(the hazard)은 상수로 가정
  - 자료의 절단없이 성공이 발생할 때까지 진행되는 시행에서 대상 n이 관측되는 시간을 $t_1, t_2, \cdots, t_n$ 이라 할 때의 추정량 $\hat{\lambda} = 1/\bar{t}$ where $\bar{t} = (1/n)\sum t_i$
    - ✓ n이 충분히 큰 경우, ML 추정이 가능하며 추정량의 분포는 $\hat{\lambda} \sim \text{app. } N\left(\lambda, \lambda^2/n\right)$

- 추정에 필요한 표본크기 결정
  - 상대허용오차 ε 이내 결과를 얻기 위해 필요한 표본크기

$$n = \left[z_{1-\alpha/2} / \varepsilon \right]^2$$

여기서 $|\hat{\lambda} - \lambda| = z_{1-\alpha/2} \cdot \lambda/\sqrt{n}$, $\varepsilon = |\hat{\lambda} - \lambda|/\lambda$

가설검정에 필요한 표본크기의 결정

- 위험함수(혹은 발병률)에 대한 가설검정의 문제는 성공이 나타나는 시간(즉, 실패시간)과 동일한 관점으로 검정이 가능 $\mu = 1/\lambda$
- 발병률 자료가 유용한 경우는 위험함수를 이용하여 직접 검정이 가능하지만, 그렇지 않은 경우는 실패시간의 평균 소요시간을 이용하여 검정 가능
- $H_0 : \lambda = \lambda_0$ 에 대한 검정 통계량 $z_0 = \sqrt{n}\left(\hat{\lambda} - \lambda_0\right)/\lambda_0 \sim N(0,1) \quad \text{under} \quad H_0$
- $H_1 : \lambda \neq \lambda_0$ 에 대한 검정력 $\Pr\left\{\sqrt{n}\left(\hat{\lambda} - \lambda_0\right)/\lambda_0 > z_{1-\alpha/2} \mid H_1\right\} = 1 - \beta$
- **양측검정을 위해 필요한 표본크기** : [참고] 모비율 추정의 그림

$$\lambda_0 + z_{1-\alpha/2}\left(\lambda_0/\sqrt{n}\right) = \lambda_1 - z_{1-\beta}\left(\lambda_1/\sqrt{n}\right)$$ 이므로

$$n = \frac{\left(z_{1-\alpha/2} \cdot \lambda_0 + z_{1-\beta} \cdot \lambda_1\right)^2}{\left(\lambda_0 - \lambda_1\right)^2}$$

Example 9

- A회사에서 특정 화학약품에 노출된 대상의 위험함수가 0.2인 데, 최근 새로운 생산 기법으로 인해 위험이 25% 정도 변화가 있었다고 가정하자. 유의수준 5%에서 수행되는 위험함수의 변화에 대한 검정결과를 80% 수준에서 신뢰할만한 결과로 얻고자 한다. 얼마나 많은 사람을 표본으로 선정해야 하는 가?

$$H_0 : \lambda = .2 \ \text{vs} \ H_1 : \lambda = 0.2 \pm (0.2 \times 0.25)$$

$$n_{0.15} = \frac{\left(z_{1-\alpha/2} \cdot \lambda_0 + z_{1-\beta} \cdot \lambda_1\right)^2}{\left(\lambda_0 - \lambda_1\right)^2} = \frac{\left(1.96 \times 0.2 + 0.842 \times 0.15\right)^2}{\left(0.2 - 0.15\right)^2} = 107.45$$

$$n_{0.25} = \frac{\left(z_{1-\alpha/2} \cdot \lambda_0 + z_{1-\beta} \cdot \lambda_1\right)^2}{\left(\lambda_0 - \lambda_1\right)^2} = \frac{\left(1.96 \times 0.2 + 0.842 \times 0.25\right)^2}{\left(0.2 - 0.25\right)^2} = 145.20$$

$$\therefore \ n = \max(n_{0.15}, n_{0.25}) = 146$$

2) Sample size determination for the incidence rate studies(2)

- $H_0: \lambda_1 = \lambda_2$ 에 대한 검정(차이의 추정보다 검정에 관심)
  - 검정통계량 $z_0 = \dfrac{(\hat{\lambda}_1 - \hat{\lambda}_2)}{\sqrt{2 \cdot \bar{\lambda}^2 / n}} = \dfrac{\sqrt{n}(\hat{\lambda}_1 - \hat{\lambda}_2)}{\bar{\lambda}\sqrt{2}}$ where $\bar{\lambda} = (\hat{\lambda}_1 + \hat{\lambda}_2) / 2$
  - $H_1: \lambda_1 \neq \lambda_2$ 에 대한 검정력 $\Pr\{z_0 > z_{1-\alpha/2} \mid H_1\} = 1 - \beta$
  - 검정을 위한 표본크기 결정

$$0 + z_{1-\alpha/2}\sqrt{2 \cdot \bar{\lambda}^2 / n} = (\lambda_1 - \lambda_2) - z_{1-\beta}\sqrt{(\lambda_1^2 + \lambda_2^2)/n}$$ 이므로

$$n = \frac{\left(z_{1-\alpha/2} \cdot \sqrt{2 \cdot \bar{\lambda}^2} + z_{1-\beta} \cdot \sqrt{\lambda_1^2 + \lambda_2^2}\right)^2}{(\lambda_1 - \lambda_2)^2}$$

  - $k = n_2 / n_1$ (불균등표본)인 경우

$$n = \frac{\left(z_{1-\alpha/2} \cdot \sqrt{(1+k) \cdot \bar{\lambda}^2} + z_{1-\beta} \cdot \sqrt{k\lambda_1^2 + \lambda_2^2}\right)^2}{k \cdot (\lambda_1 - \lambda_2)^2}$$

여기서 $\bar{\lambda} = (\hat{\lambda}_1 + k\hat{\lambda}_2) / (1+k)$

Example 10

- B회사에서 특정 화학약품에 노출된 경우 사망 위험이 약 0.1이며, 경쟁사의 사망 위험이 0.05라고 하자. 유의수준 5%에서 두 회사의 사망 위험에 차이가 있는지 검정을 하고자 하며, 검정 결과에 대한 신뢰정도를 80%수준에서 유지하고자 한다. 이 연구를 위해 실시되는 추적조사(follow-up survey)에 필요한 표본크기는 각각 얼마인가?

$$\bar{\lambda} = {(\lambda_1 + \lambda_2)}\big/{2} = 0.075$$

$$n = \frac{\left(z_{1-\alpha/2} \cdot \sqrt{2 \cdot \bar{\lambda}^2} + z_{1-\beta} \cdot \sqrt{\lambda_1^2 + \lambda_2^2}\right)^2}{(\lambda_1 - \lambda_2)^2} = \frac{\left(1.96\sqrt{2 \times 0.075^2} + 0.842\sqrt{0.1^2 + 0.05^2}\right)^2}{(0.1 - 0.05)^2} = 36.49$$

3) Sample size determination for the incidence rate studies(3)

- 연구기간을 설정한 후 실시되는 추적조사에 필요한 표본크기의 결정
  - Gross and Clark(1975)의 연구 : 총 연구기간 T 로 설정

$$n = \frac{\left(z_{1-\alpha/2} \cdot \sqrt{2 \cdot f(\bar{\lambda})} + z_{1-\beta} \cdot \sqrt{f(\lambda_1) + f(\lambda_2)}\right)^2}{(\lambda_1 - \lambda_2)^2}$$

$$\text{where} \quad f(\lambda) = \lambda^3 \cdot T \Big/ \left(\lambda \cdot T - 1 + e^{-\lambda \cdot T}\right)$$

  - Lachin(1981)의 연구 : 최초연구기간 T_1 , 총 연구기간은 T 로 설정

$$n = \frac{\left(z_{1-\alpha/2} \cdot \sqrt{2 \cdot g(\bar{\lambda})} + z_{1-\beta} \cdot \sqrt{g(\lambda_1) + g(\lambda_2)}\right)^2}{(\lambda_1 - \lambda_2)^2}$$

$$\text{where} \quad g(\lambda) = \lambda^3 \cdot T_1 \Big/ \left(\lambda \cdot T_1 - e^{-\lambda \cdot (T - T_1)} + e^{-\lambda \cdot T}\right)$$

Example 11

- 예제 10의 연구를 5년 동안 진행할 예정이라고 한다. 연구자는 $\alpha = 0.05$ and $\beta = 0.2$ 의 수준에서 가설 $H_0 : \lambda_1 = \lambda_2 = 0.1$ 에 대한 $H_1 : \lambda_1 = 0.1$ and $\lambda_2 = 0.05$ 의 가설을 검정하려고 한다. 추적조사 연구에 필요한 표본규모는 얼마인가?

$$f(\bar{\lambda}) = \bar{\lambda}^3 \cdot T \Big/ \left(\bar{\lambda} \cdot T - 1 + e^{-\bar{\lambda} \cdot T}\right) = 0.0339 \quad , \quad f(\lambda_1) = 0.0469 \ , \quad f(\lambda_2) = 0.0217$$

$$\therefore \quad n = \frac{\left(z_{1-\alpha/2} \cdot \sqrt{2 \cdot f(\bar{\lambda})} + z_{1-\beta} \cdot \sqrt{f(\lambda_1) + f(\lambda_2)}\right)^2}{(\lambda_1 - \lambda_2)^2}$$

$$= \frac{\left(1.96\sqrt{2 \times 0.0339} + 0.842\sqrt{(0.0469 + 0.0217)}\right)^2}{(0.1 - 0.05)^2} = 213.7$$

Example 12

- 어느 질병에 대해 표준의 처치를 받은 환자의 평균 생존기간은 2년이라고 한다. 새로운 처치가 개발되어 평균 생존기간이 최소 1년이 더 연장되었다고 주장한다. 새로운 처치에 대한 연구를 5년 동안 진행하면서 새로운 처치에 대한 치료 결과를 검정하려고 한다. 이 연구에 필요한 표본규모는 각각 얼마인가?

$$\lambda_1 = 1/\mu_1 = 0.5,\quad \lambda_2 = 1/\mu_2 = 0.333,\quad \bar{\lambda} = 0.4167$$

$$f(\bar{\lambda}) = \bar{\lambda}^3 \cdot T \Big/ \left(\bar{\lambda} \cdot T - 1 + e^{-\bar{\lambda} \cdot T}\right) = 0.2995,\quad f(\lambda_1) = 0.3950,\quad f(\lambda_2) = 0.2164$$

$$\therefore\quad n = \frac{\left(z_{1-\alpha/2} \cdot \sqrt{2 \cdot f(\bar{\lambda})} + z_{1-\beta} \cdot \sqrt{f(\lambda_1) + f(\lambda_2)}\right)^2}{(\lambda_1 - \lambda_2)^2}$$

$$= \frac{\left(1.645\sqrt{2 \times 0.2995} + 0.842\sqrt{(0.3950 + 0.2164)}\right)^2}{(0.5 - 0.333)^2} = 163.74$$

Example 13

- 예제 12의 연구를 전/후반기로 나뉘어 전반기에 처음 2.5년 동안 진행한 후 후반기 연구로 2.5년을 계속 진행하려고 계획하고 있다. 이 추적조사 연구에 필요한 표본규모는 각각 얼마인가? (표본크기는 전반기에 필요한 표본규모를 우선 결정한 후 후반기에는 전반기 결과를 반영하여 표본규모를 정하게 됨)

$$g(\bar{\lambda}) = \bar{\lambda}^3 \cdot T_1 \Big/ \left(\bar{\lambda} \cdot T_1 - e^{-\bar{\lambda}\cdot(T - T_1)} + e^{-\bar{\lambda}\cdot T}\right) = 0.2224\ ,\ \ g(\lambda_1) = 0.2989\ ,\ \ g(\lambda_2) = 0.1576$$

$$\therefore\quad \mathrm{n} = \frac{\left(\mathrm{z}_{1-\alpha/2} \cdot \sqrt{2 \cdot g(\bar{\lambda})} + \mathrm{z}_{1-\beta} \cdot \sqrt{g(\lambda_1) + g(\lambda_2)}\right)^2}{(\lambda_1 - \lambda_2)^2}$$

$$= \frac{\left(1.645\sqrt{2 \times 0.2224} + 0.842\sqrt{0.2989 + 0.1576}\right)^2}{(0.5 - 0.3333)^2} = 99.88$$

Sample size
for continuous response variables

- 모평균 추정의 경우

$$n = \frac{z^2_{1-\alpha/2}\ \sigma^2}{d^2} \quad \text{where} \quad d = z_{1-\alpha/2}\sqrt{\sigma^2 / n}$$

  - 상대오차로 주어진 경우

$$n = \frac{z^2_{1-\alpha/2}\ \sigma^2}{\varepsilon^2 \mu^2} \quad \text{where} \quad \varepsilon = \frac{|\hat{\mu} - \mu|}{\mu}$$

- 가설검정 $H_1 : \mu > \mu_0$ 에 필요한 표본크기의 결정

$$n = \frac{\sigma^2 \left(z_{1-\alpha} + z_{1-\beta}\right)^2}{\left(\mu_0 - \mu_1\right)^2}$$

□ 두 모평균 차이 추정에 필요한 표본크기의 결정

$$\mathrm{Var}\,(\overline{X}_1 - \overline{X}_2) = \sigma_1^2 / n_1 + \sigma_2^2 / n_2 \Rightarrow \sigma^2 \left(1/n_1 + 1/n_2 \right) \Rightarrow 2 \cdot \sigma^2 / n$$

$$n = \frac{z_{1-\alpha/2}^2 \left(2 \cdot \sigma^2\right)}{\mathrm{d}^2} \quad \text{where} \quad \mathrm{d} = z_{1-\alpha/2}\sqrt{2\sigma^2 / n}$$

□ 가설검정 $H_1 : \mu_1 > \mu_2$ 에 필요한 표본크기의 결정

$$n = \frac{2\sigma^2 \left(z_{1-\alpha} + z_{1-\beta}\right)^2}{\left(\mu_1 - \mu_2\right)^2} \qquad \text{cf.} \quad s_P^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{(n_1 - 1) + (n_2 - 1)}$$